(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 387 319 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
06.03.1996 Bulletin 1996/10

(21) Application number: **89909015.3**

(22) Date of filing: **14.07.1989**

(51) Int Cl.[6]: **C12N 15/81**
// (C12N15/81, C12R1:865)

(86) International application number:
PCT/GB89/00816

(87) International publication number:
WO 90/01063 (08.02.1990 Gazette 1990/04)

(54) **SECRETORY LEADER SEQUENCES**

SEKRETORISCHE LEADER-SEQUENZEN

SEQUENCES CONDUCTRICES SECRETOIRES

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(30) Priority: **23.07.1988 GB 8817598**
28.03.1989 GB 8906920

(43) Date of publication of application:
19.09.1990 Bulletin 1990/38

(73) Proprietor: **Delta Biotechnology Limited**
Nottingham NG7 1FD (GB)

(72) Inventors:
- **GOODEY, Andrew, Robert**
Mapperley Park Nottingham (GB)
- **BELFIELD, Graham, Paul**
Beeston Nottingham (GB)
- **SLEEP, Darrell**
West Bridgford Nottingham (GB)

(74) Representative: **Bassett, Richard Simon**
ERIC POTTER & CLARKSON
St. Mary's Court
St. Mary's Gate
Nottingham NG1 1LE (GB)

(56) References cited:
EP-A- 0 123 544
EP-A- 0 127 304
EP-A- 0 206 783
EP-A- 0 220 689
EP-A- 0 252 561

- **BIO/TECHNOLOGY (1990) 8:42-46, D. Sleep et al.**
- **BIO/TECHNOLOGY (1989) 7:55-60, Stetler et al.**
- **GENE (1990) 90:255-262**
- **INT. J. PEPTIDE PROTEIN RES. (1991) 38:401-408, Craig et al.**
- **Dialog Information Services, File 351, World Patent Index 81-89, Dialog accession no. 88-195319/28, Agency of Ind Sci Tech**

Remarks:

The file contains technical information submitted after the application was filed and not included in this specification

EP 0 387 319 B1

Printed by Jouve, 75001 PARIS (FR)

Description

New secretory leader sequences.

This invention relates to secretory leader sequences which can be employed to direct the secretion of a heterologous protein (such as human serum albumin) from fungi (for example the yeast Saccharomyces cerevisiae).

Translocation of protein molecules through bi-lipid membranes from one cellular compartment to another generally relies upon information held within the primary amino acid sequence of the protein itself. The most prevalent and therefore the best characterised sequence information is the amino terminal leader or signal sequence of prokaryotic and eukaryotic organisms. Genetic studies in which the signal sequence has been totally or extensively deleted indicate that the signal sequence is essential for protein translocation (Benson, S.A. et al. 1985, Ann. Rev. Biochem. 54, 101-134). Among several hundred known sequences (Watson, M.E.E., 1984, Nuc. Acid. Res. 12, 5145-5164) no consensus signal sequence or even an absolute requirement for any amino acid at any given position can be discerned, although a common feature of many leader sequences is a core of 7-10 hydrophobic amino acids. Genetic manipulations which result in alterations to the hydrophobic core, either by deletion or by inserting charged residues, generally result in a block in protein translocation (Benson, S.A., et al. 1985, Ann. Rev. Biochem. 54, 101-134). Moreover, in a series of extensive modifications to the chicken lysozyme leader sequence, Yamamoto et al. 1987 (Biochem. and Biophys. Res. Comm. 149, 431-436) have shown that, while some alterations to the hydrophobic core can result in the abolition of secretion, others can potentiate the leader sequence function, resulting in increased levels of protein secretion.

While the leader sequence is usually essential for the translocation of proteins across membranes, once translocated these sequences are usually endoproteolytically cleaved by enzymes contained within the cellular compartments into which the proteins have now moved. These enzymes recognise specific amino acid sequences within the primary structure of the translocated protein. Moreover, complete processing of certain eukaryotic proteins to their mature form often relies upon a series of proteolytic cleavages (Bussey, H., 1988 Yeast 4, 17-26).

With the recent advances in recombinant DNA technology, increasing resources have been brought to bear on the commercial exploition of fungi, particularly yeasts, as vehicles for the production of a diverse range of proteins.

Since many of these proteins are themselves naturally secreted products, it is possible to utilise the information contained within the leader sequence to direct the protein through the secretion pathway. However, this information is contained within a peptide foreign to yeast. Its recognition and subsequent processing by the yeast secretory pathway are not necessarily as efficient as those of a homologous yeast leader sequence. As a consequence an alternative approach has been to replace the leader sequence with one derived from a naturally secreted yeast protein.

The most widely used yeast secretory sequence is the 89 amino acid leader sequence of the alpha-factor mating pheromone. Processing of this leader has been extensively studied (Kurjan & Herskowitz, Cell 30, 933-943, 1982; Julius et al. 1983 Cell 32, 839-852; Dmochowska et al. Cell 50, 573-584, 1987; Julius et al. Cell 36: 309-318, 1984; Julius et al. Cell 37, 1075-1085, 1984) and requires at least four gene products for complete proteolytic cleavage to liberate the mature 13 amino acid alpha-factor pheromone.

Complete proteolytic cleavage of the alpha-factor primary translation product requires first the removal of the N-terminal 19 amino acid signal sequence by a signal peptidase within the endoplasmic reticulum. Following this the sequential action of three gene products located within the golgi apparatus processes the large precursor molecule, liberating four copies of the alpha-factor pheromone. These are the KEX2 gene product, an endopeptidase which cleaves after the Lys-Arg dibasic amino acid pair, a carboxypeptidase β-like cleavage, recently identified as the product of the KEX1 gene, and a dipeptidyl amino peptidase, the product of the STE13 gene, which sequentially removes the Glu-Ala or Asp-Ala diamino acid pairing preceding the mature alpha-factor pheromone.

The alpha factor prepro leader sequence has successfully been employed to secrete a range of diverse proteins and peptides. However, when the alpha-factor signal is used to direct secretion of human serum albumin, we have found that a large proportion of the extracellular HSA produced is in the form of a 45KD N-terminal fragment.

EP-A-252 561 (Sclavo) discloses the use of the 16 amino acid signal peptide (pre-sequence) from the killer toxin of Kluyveromyces lactis to aid secretion of heterologous proteins in yeast.

A further possibility is to use a fusion secretory leader sequence. This may be generated by the fusion of two independent sequences. A hybrid signal in which the first amino acids of the acid phosphatase signal were fused to the proteolytic cleavage site of human alpha interferon resulted in the expression and secretion of interferon (Hinnen et al. Foundation for Biochemical and Industrial Fermentation Research, 229, 1219-1224, 1983); 10% of the interferon produced was secreted into the medium. In a similar approach the first 22 amino acids of the alpha-factor leader were fused to the last twelve amino acids of the human interferon alpha-2 signal sequence resulting in the secretion of interferon alpha-2 into the culture supernatant (Piggott et al. Curr. Genet. 12 561-567, 1987). An identical construct in which the interferon alpha-2 gene was replaced by the interferon β gene did not result in any secretion of human interferon β into the culture supernatant. Finally, in a series of experiments designed to assess the effect of leader sequences on the secretion of human lysozyme, Yoshimura et al. (Biochem. & Biophys. Res. Comm. 145, 712-718, 1987) described a

fusion leader comprising the first 9 amino acids of the chicken lysozyme leader and the last 9 amino acids of the Aspergillus awamori glycoamylase leader. Although this fusion leader was effective in secreting 60% of the produced material into the culture supernatant, it was only 15% as effective as the entire chicken lysozyme leader. Moreover, no secreted product could be detected if the human lysozyme sequences were preceded by the entire Aspergillus glycoamylase leader, or a fusion derived from the first 9 amino acids of the Aspergillus glucoamylase leader and the last 9 amino acids of the chicken lysozyme leader.

EP-A-0 220 689 disclosed fusions of the S. cerevisiae barrier protein leader sequence and part of the barrier protein with proinsulin.

We have now devised new and advantageous leader sequences for use in fungi.

One aspect of the invention provides an amino acid sequence as follows:

(a) H_2N-Met-Lys-Trp-Val-Ser-Phe-Ile-Ser-Leu-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-COOH

or

(b) H_2N-Met-Asn-Ile-Phe-Tyr-Ile-Phe-Leu-Phe-Leu-Leu-Ser-Phe-Val-Gln-Gly-Ser-Leu-Asp-Lys-Arg-COOH

or conservatively modified variations of either sequence, defined as follows.

Table 1 shows alternative amino acids for each position except the initial methionine. Any of the possible permutations are within the scope of the invention. The selection of lysine or arginine for the last two positions is particularly non-critical, although there should always be Lys or Arg at each of these positions. Positions 20 and 21 of sequence (a) are not Gly and Val respectively. Sequences which are up to four amino acids shorter or longer are also included provided that the C-terminal Lys-Arg, Arg-Lys, Lys-Lys or Arg-Arg entity is maintained, there is a positively charged residue within 5 residues of the N-terminus and there is a generally hydrophobic region at or adjacent the middle of the sequence. The pentapeptide adjacent the C-terminal is retained.

Table 1

Leader (a)

1								10			
Met	Lys	Trp	Val	Ser	Phe	Ile	Ser	Leu	Leu	Phe	Leu	Phe	Ser
	Arg	Phe	Leu	Thr	Trp	Leu	Thr	Ile	Ile	Trp	Ile	Trp	Thr
	His	Tyr	Ile	Gly	Tyr	Val	Gly	Val	Val	Tyr	Val	Tyr	Gly
	Gln		Met	Ala		Met	Ala	Met	Met		Met		Ala
	Asn												

				20					
Ser	Ala	Tyr	Ser	Arg	Ser	Leu	Asp	Lys	Arg
Thr	Thr	Phe	Thr	Lys	Thr	Ile	Glu	Arg	Lys
Gly	Gly	Trp	Gly	His	Gly	Val	Asn		
Ala	Ser		Ala	Gln	Ala	Met	Gln		
				Asn			His		

Leader (b)

```
Met Asn Ile Phe Tyr Ile Phe Leu Phe Leu Leu Ser Phe Val
    Asp Leu Trp Phe Leu Trp Ile Trp Ile Ile Thr Trp Leu
    Glu Val Tyr Trp Val Tyr Val Tyr Val Val Gly Tyr Ile
    Gln Met         Met     Met     Met Met Ala     Met
    His
```

```
Gln Gly Ser Leu Asp Lys Arg
Asp Ser Thr Ile Asn Arg Lys
Asn Thr Gly Val Glu
Glu Ala Ala Met Gln
His             His
```

A second aspect provides a fusion compound comprising any of the said amino acid sequences linked, preferably directly, at the carboxyl terminal to the N-terminal residue of a polypeptide. The polypeptide may be any desired polypeptide, including "pro-polypeptides" (in other words precursors which undergo post-translational cleavage or other modification, such as glycosylation). The term "polypeptide" encompasses oligopeptides. The polypeptide may be fibronectin or a portion thereof (for example the collagen or fibrin-binding portions described in EP 207 751), urokinase, pro-urokinase, the 1-368 portion of CD4 (D. Smith et al (1987) Science 328, 1704-1707), platelet derived growth factor (Collins et al (1985) Nature 316, 748-750), transforming growth factor β (Derynck et al (1985) Nature 316, 701-705), the 1-272 portion of Von Willebrand's Factor (Bontham et al, Nucl. Acids Res. 14 7125-7127), the Cathepsin D fragment of fibronectin (585-1578), α_1-antitrypsin, plasminogen activator inhibitors, factor VIII, α-globin, β-globin, myoglobin or nerve growth factor or a conservative variant of any of these. The polypeptide may also be a fusion of HSA or an N-terminal portion thereof and any other polypeptide, such as those listed above. Preferably, the polypeptide is a naturally-occurring human serum albumin, a modified human serum albumin or a fragment of either, such modified forms and fragments being termed "variants". These variants include all forms or fragments of HSA which fulfill at least one of the physiological functions of HSA and which are sufficiently similar to HSA, in terms of structure (particularly tertiary structure) as to be regarded by the skilled man as forms or fragments of HSA.

In particular variants or fragments of HSA which retain at least 50% of its ligand-binding properties, for example with respect to bilirubin or fatty acids, (preferably 80%, or 95%) are encompassed. Such properties are discussed in Brown, J.R. & Shockley, P. (1982) in Lipid-Protein Interactions 1, 26-68, Ed. Jost, P.C. & Griffith, O.H.

The portion of HSA disclosed in EP 322 094 is an example of a useful fragment of HSA which may be secreted by use of the leader sequences of the invention.

A third aspect provides a nucleotide sequence coding for any of the said amino acid sequences or for the said fusion compound. The nucleotide sequence (or the portion thereof encoding the leader sequence) may be selected from the possibilities shown in Tables 2 & 3, for sequences (a) and (b) respectively, where the codons encoding each amino acid are listed under the amino acids. The codons of Tables 2 and 3 clearly relate to RNA, but it is to be understood that equivalent DNA nucleotide sequences are also within the scope of this aspect of the invention.

Table 2

Met	Lys	Trp	Val	Ser	Phe	Ile	Ser	Leu	Leu	Phe	Leu	Phe	Ser
AUG	AAA	UGG	GUU	UCU	UUU	AUU	UCU	UUA	UUA	UUU	UUA	UUU	UCU
	AAG		GUC	UCC	UUC	AUC	UCC	UUG	UUG	UUC	UUG	UUC	UCC
			GUA	UCA		AUA	UCA	CUU	CUU		CUU		UCA
			GUG	UCG			UCG	CUC	CUC		CUC		UCG
				AGU			AGU	CUA	CUA		CUA		AGU
				AGC			AGC	CUG	CUG		CUG		AGC

Ser	Ala	Tyr	Ser	Arg	Ser	Leu	Asp	Lys	Arg
UCU	GCU	UAU	UCU	CGU	UCU	UUA	GAU	AAA	CGU
UCC	GCC	UAC	UCC	CGC	UCC	UUG	GAC	AAG	CGC
UCA	GCA		UCA	CGA	UCA	CUU			CGA
UCG	GCG		UCG	CGG	UCG	CUC			CGG
AGU			AGU	AGA	AGU	CUA			AGA
AGC			AGC	AGG	AGC	CUG			AGG

Table 3

Met	Asn	Ile	Phe	Tyr	Ile	Phe	Leu	Phe	Leu	Leu	Ser	Phe	Val
AUG	AAU	AUU	UUU	UAU	AUU	UUU	UUA	UUU	UUA	UUA	UCU	UUU	GUU
	AAC	AUC	UUC	UAC	AUC	UUC	UUG	UUC	UUG	UUG	UCC	UUC	GUC
		AUA			AUA		CUU		CUU	CUU	UCA		GUA
							CUC		CUC	CUC	UCG		GUG
							CUA		CUA	CUA	AGU		
							CUG		CUG	CUG	AGC		

Gln	Gly	Ser	Leu	Asp	Lys	Arg
CAA	GGU	UCU	UUA	GAU	AAA	CGU
CAG	GGC	UCC	UUG	GAC	AAG	CGC
	GGA	UCA	CUU			CGA
	GGG	UCG	CUC			CGG
		AGU	CUA			AGA
		AGC	CUG			AGG

A fourth aspect provides a DNA construct comprising a suitable control region or regions and a nucleotide sequence as defined above, the sequence being under the control of the control region. By "suitable control region" we mean such DNA regions as are necessary to enable the said nucleotide sequence to be expressed in the host for which the construct is intended. The control region will usually include transcriptional start and stop sequences, 3'-polyadenylation sequences, a promoter and, often, an upstream activation site for the promoter. The man skilled in the art will readily be able to select and assemble suitable regions from those available in this art. However, specific examples of suitable expression vectors and their construction include those disclosed in EP 198 745, GB 2 171 703 (for B.subtilis), EP 207 165, EP 116 201, EP 123 244, EP 123 544, EP 147 198, EP 201 239, EP 248 637, EP 251 744, EP 258 067, EP 286 424 and EP 322 094.

A fifth aspect provides a host transformed with the said DNA construct. The host may be any host in which the construct is found to work adequately, including bacteria, yeasts, filamentous fungi, insect cells, plant cells and animal cells. Preferably, however, the host is Saccharomyces cerevisiae or Schizosaccharomyces pombe, most preferably the former. As many native secretion signals are effective in heterologous hosts (for example the natural HSA leader sequence in yeast) it is entirely reasonable to suppose that the leader sequences of the invention will function in hosts other than yeasts.

A sixth aspect provides a process for preparing a polypeptide, comprising cultivating the said host and obtaining therefrom the polypeptide expressed by the said nucleotide sequence, or a modified version thereof.

By "modified version thereof", we mean that the actual polypeptide which is separated may have been post-translationally modified, in particular by cleavage of the leader sequence.

So that the invention may be more readily understood, preferred aspects will now be illustrated by way of example and with reference to the accompanying drawings in which:

Figure 1 is a restriction map of plasmid pEK113;

Figure 2 is a restriction map of plasmid pEK25;

Figure 3 is a restriction map of plasmid pAYE230;

Figure 4 is a restriction map of plasmid pAYE238;

Figure 5 is a restriction map of plasmid pAYE304; and

Figure 6 is a restriction map of plasmid pAYE305.

Example of a prior art type of leader sequence

The DNA coding sequence for mature HSA protein has been placed immediately downstream of a DNA sequence encoding the KEX2 cleavage site of the alpha factor pre pro leader sequence (85 amino acids). When this protein sequence is placed under the control of a promoter on a yeast autonomously replicating plasmid and transformed into a haploid strain of the yeast Saccharomyces cerevisiae, mature HSA can be detected in the culture supernatant. N-terminal amino acid sequence information indicates that the secreted protein has the same N-terminal amino acid composition as natural HSA, namely Asp-Ala-His. This also indicates that the first two amino acids of the secreted HSA are not susceptible to the dipeptidyl endopeptidase, the product of the STE13 gene, as this enzyme is responsible for the removal of such sequences from between successive repeats of the alpha-factor pheromone. Although mature HSA is the major product observed in the culture supernatant, a N-terminal fragment of HSA (45 kilodaltons) was also detected, representing approximately 15% of the total HSA synthesised. This fragment component represents not only a waste of secretion capacity but also certain downstream purification problems in that, as a fragment of HSA, it shares some biochemical and biophysical properties with intact HSA.

EXAMPLE 1

We have constructed a fusion leader which may be regarded as the natural HSA leader sequence from which the last five amino acids have been removed, to be replaced by the five amino acids preceding the KEX2 cleavage site of the alpha-factor pre pro leader sequence, i.e. amino acids 81 to 85, Ser-Leu-Asp-Lys-Arg (Table 2).

When transformed with suitable plasmid vectors incorporating the fusion leader, yeast secrete mature HSA into the culture supernatant at levels comparable to that observed with the alpha-factor leader sequence. N-terminal sequence analysis indicates that the mature HSA possesses the correct N-terminal amino acid composition.

Moreover, substitution of the alpha-factor leader by the fusion leader sequence has been found to result in a 6 fold reduction in the levels of the 45 kd fragment observed in the culture supernatant. This therefore represents a significant improvement in the reduction of the contaminating polypeptides, thus aiding the purification of mature HSA from yeast culture supernatants.

Details

Unless otherwise stated all procedures were carried out as described by Maniatis et al (1982). Plasmid pEK113 (Figure 1) (EP-A-248 637) was digested to completion with the restriction endonucleases MstII and HindIII. DNA was recovered by phenol/chloroform extraction and ethanol precipitation. The linearised plasmid DNA was then treated with the Klenow fragment of E.coli DNA polymerase I to generate a linearised DNA molecule with blunt ends.

The following oligonucleotide duplex (I) was constructed on an automated Applied Biosystems Inc 380B DNA synthesiser (according to manufacturer's instructions).

Oligonucleotide I

5′ 3′

GGC TTA TAA GGA TCC TTA TAA GCC

CCG AAT ATT CCT AGG AAT ATT CGG

The oligonucleotide duplex was ligated with equimolar quantities of linearised, blunt ended pEK113. E.coli strain

MC1061 was transformed with the ligation mixture and cells receiving DNA were selected on an ampicillin-containing medium (50ug/ml ampicillin). Recombinant plasmids containing the oligonucleotide duplex were screened by digesting DNA prepared from individual colonies with the restriction endonucleases MstII and EcoRI. Plasmid pEK25 was thus formed (Figure 2).

Plasmid pEK25 was digested to completion with the restriction endonucleases XbaI and BamHI, DNA fragments were separated by electrophoresis through a 1% (w/v) agarose gel and a 688 base pair XbaI - BamHI DNA fragment recovered from the gel by electroelution.

The plasmid mp19.7 (EP-A-248 637) was digested to completion with the restriction endonuclease XhoI. The linearised DNA was phenol/chloroform extracted and ethanol precipitated. The recovered DNA was then treated with the Klenow fragment of E. coli DNA polymerase I as previously described, following which the DNA was phenol/chloroform extracted and ethanol precipitated. The recovered DNA was then digested to completion with XbaI and the digestion products separated by agarose gel electrophoresis. A 1067 base pair fragment was recovered from the gel by electroelution. The following oligonucleotide duplex (II) was prepared as described previously.

Oligonucleotide II

```
5'
  GATCC ATG AAG TGG GTA AGC TTT ATT TCC CTT CTT TTT CTC
        TAC TTC ACC CAT TCG AAA TAA AGG GAA GAA AAA GAG

                                                    3'
  TTT AGC TCG GCT TAT TCC AGG AGC TTG GAT AAA AGA
  AAA TCG AGC CGA ATA AGG TCC TCG AAC CTA TTT TCT
```

The plasmid pUC19 (Yanisch-Perron et al. 1985) was digested to completion with the restriction endonuclease BamHI. Linearised DNA was recovered by phenol/chloroform extraction and ethanol precipitation.

Equimolar quantities of the BamHI digested pUC19, the oligonucleotide duplex II, the 1067 b.p. DNA fragment derived from mp19.7 and the 688 b.p. DNA fragment derived from pEK25 were ligated together. E.coli DH5 was transformed with the ligated DNA and transformants selected on 50ug/ml ampicillin L-broth agar. Recombinant colonies containing the desired plasmid, designated pAYE 230 (Figure 3) were selected by digested DNA obtained from individual colonies with the restriction endonuclease BamHI.

Plasmid pAYE 230 was digested to completion with BamHI and the products separated by electrophoresis through a 1% agarose gel. The 1832 base pair fragment containing the HSA coding sequence was recovered by electroelution.

Plasmid pMA91 (Mellor et al. 1983) was digested to completion with BglII under standard conditions. The linearised plasmid was phenol/chloroform extracted and ethanol precipitated.

Equivalent quantities of the linearised pMA91 and the DNA fragment prepared from pAYE 230 were ligated under standard conditions. E. coli DH5 was transformed with the ligation mixture and cells receiving the DNA selected on L-broth agar containing 50µg/ml ampicillin. Colonies containing the desired plasmid, designating pAYE 238 (Figure 4) were selected by digesting the DNA from such colonies with PvuII.

Plasmid pAYE 238 was transformed into the yeast Saccharomyces cerevisiae strain S150-2B as described by Hinnen et al. (1978). Cells receiving plasmid pAYE 238 were selected on minimal medium, supplemented with 2% (w/v) glucose, 20mg/l histidine, 20mg/l tryptophan and 20mg/l uracil.

Transformed S150-2B cells were transferred to 10ml YEPD media containing 2% (w/v) glucose and incubated at 30°C, 200rpm for 72 hours. Cell free culture supernatants were analysed by discontinuous native 8-25% gradient polyacrylamide gel electrophoresis on a Pharmacia Phast System, as described in the manufacturer's instructions. Cells were stained and destained and the relative quantities of native HSA and HSA fragment estimated by gel scan at 595nm.

EXAMPLE 2

We have also constructed a second fusion leader which consists of the 16 amino acid pre region of the 97,000 dalton Kluyveromyces lactis killer (ORF 2) toxin (Stark and Boyd, 1986, Tokumaga et al 1987) fused to the five amino acids preceding the KEX2 cleavage site of the alpha-factor prepro leader sequence, i.e. amino acids 81 to 85, Ser-Leu-Asp-Lys-Arg (Table 3).

When transformed with plasmid vectors incorporating the fusion leader described in Table 3, yeast secreted mature HSA into the culture supernatants at levels higher than when either the natural K.lactis prepro killer toxin leader sequence or the alpha-factor prepro leader sequence was used. N-terminal sequence analysis indicates that the mature HSA possesses the correct N-terminal amino acid composition.

Substitution of the alpha-factor leader by the K.lactis killer/alpha factor fusion leader sequence resulted in a six fold reduction in the levels of the 45kd fragment observed in the culture supernatant. This therefore represents a significant improvement in the reduction of the contaminating polypeptides, thus aiding the purification of mature HSA from yeast culture supernatants.

Details

The experimental procedures employed to generate a yeast HSA secretion vector utilising the K.lactis killer/alpha factor fusion leader were identical to those described in Example 1, except that oligonucleotide duplex (II) was replaced by oligonucleotide duplex (III) synthesised on an automated Applied Biosystems Inc. 380B DNA synthesiser (according to manufacturer's instructions).

Oligonucleotide duplex III

```
GATCC ATG AAT ATA TTT TAC ATA TTT TTG TTT TTG CTG TCA TTC
      TAC TTA TAT AAA ATG TAT AAA AAC AAA AAC GAC AGT AAG

          GTT CAA GGA AGC TTG GAT AAA AGA
          CAA GTT CCT TCG AAC CTA TTT TCT
```

Equimolar quantities of the BamHI digested pUC19, the oligonucleotide duplex III, the 1067bp DNA fragment derived from mp19.7 and the 688b.p. DNA fragment derived from pEK25 were ligated together. E.coli DH5 was transformed with ligated DNA and transformants selected on 50µg/ml ampicillin L-broth agar. Recombinant colonies containing the desired plasmid, designated pAYE304 (Figure 5), were selected by digested DNA obtained from individual colonies with the restriction endonuclease BamHI.

Plasmid pAYE304 was digested to completion with BamHI and the products separated by electrophoresis through a 1% agarose gel. The 1823 base pair fragment containing the HSA coding sequence was recovered by electroelution.

Plasmid pMA91 (Mellor et al, 1983) was digested to completion with BglII under standard conditions. The linearised plasmid was phenol/chloroform extracted and ethanol precipitated.

Equivalent quantities of the linearised pMA91 and the DNA fragment prepared from pAYE304 were ligated under standard conditions. E.coli DH5 was transformed with the ligation mixture and cells receiving DNA selected on L-broth agar containing 50µg/ml ampicillin. Colonies containing the desired plasmid, designating pAYE305 (Figure 6), were selected by digesting the DNA from such colonies with PvuII.

Plasmid pAYE305 was transformed into the yeast Saccharomyces cerevisiae strain S150-2B as described by Hinnen et al, (1978). Cells receiving plasmid pAYE305 were selected on minimal medium, supplemented with 2% (w/v) glucose, 20mg/l histidine, 20mg/l tryptophan and 20mg/l uracil.

Transformed S150-2B cells were transferred to 10ml YEPD medium containing 2% (w/v) glucose and incubated at 30°C, 200rpm for 72 hours. Cell free culture supernatants were analysed by discontinuous native 8-25% gradient polyacrylamide gel electrophoresis on a Pharmacia Phast System, as described in the manufacturer's instructions.

Cells were stained and destained and the relative quantities of native HSA and HSA fragment estimated by gel scan at 595nm.

EXAMPLE 3

Using a vector based on the disintegration vectors of EP286424 (Delta Biotechnology), a suitable promoter and the fusion leader of Example 1 above, Schizosaccharomyces pombe (strain Leu1.32h) was transformed and fermented at 30°C in 10ml of EMM (Edinburgh minimal medium, Ogden, J.E. & Fantes, P.A. (1986) Curr. Genetics 10 509-514), buffered to pH 5.6 with 0.1M citric acid/sodium phosphate, to give 10-15 mg/l of HSA in the culture supernatant after 3 days.

References

Beggs, J. D. (1978) Nature 275, 104-109.

Beggs, J. D. (1981), Molecular Genetics in Yeast, Alfred Benzon Symp. 16, 383-395.

Birnboim, H. C. & Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523.

Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.

Henderson, R. C. A., Cox, B. S. & Tubb, R. (1985) Curr. Genet. 9, 135-136.

Hinnen et al. (1988) PNAS 75: 1929.

Hitzeman, R. A., Clarke, L. & Carbon, J. (1980) J. Biol. Chem., 255(24), 12073-12080.

Hitzeman, R. A., Hagie, F. E., Levine, H. L., Goeddel, D. V., Ammerer, G. & Hall, B. D. (1981) Nature 293, 717-722.

Julius, D., Brake, A., Blair, L., Kunisawa, R. & Thorner, J. (1984) Cell 37, 1075-1089.

Maniatis et al. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y.

Mandel, M. & Higa, A. (1970) J. Mol. Biol. 53, 159-162.

Mellor et al. (1983) Gene 24, 1-14.

Sherman, F., Fink, G. R. & Lawrence, C.(1979) Methods in Yeast Genetics, Cold Spring Harbor, N.Y.

Sleep, D., Belfield, G. P. & Goodey, A. R. (1988) Yeast 4 (14th Int. Conf. on Yeast Genet. & Mol. Biol., Helsinki, Conf. Poster).

Stark, M. J. R. & Boyd, A. (1986) E.M.B.O.J. 5, 1995-2002.

Tokumaga, M., Wada, N. & Hishinuma, F. (1987) Biochem. Biophys. Res. Comm. 144, 613-619.

Towbin, H., Staehelin, T. & Gordon, J. (1979). P.N.A.S. 76, 4350-4354.

Vogelstein, B. (1987), Anal. Biochem. 160, 115-118.

Yanisch-Perron et al. (1985) Gene 33 103-109.

Claims

1. A polypeptide having an amino acid sequence as follows:

(a)

```
1                                                   10
Met Lys Trp Val Ser Phe Ile Ser Leu Leu Phe Leu Phe Ser
    Arg Phe Leu Thr Trp Leu Thr Ile Ile Trp Ile Trp Thr
    His Tyr Ile Gly Tyr Val Gly Val Val Tyr Val Tyr Gly
    Gln     Met Ala     Met Ala Met Met     Met     Ala
    Asn

                        20
Ser Ala Tyr Ser Arg Ser Leu Asp Lys Arg
Thr Thr Phe Thr Lys Thr Ile Glu Arg Lys
Gly Gly Trp Gly His Gly Val Asn
Ala Ser     Ala Gln Ala Met Gln
                Asn         His
```

or

(b)

```
Met Asn Ile Phe Tyr Ile Phe Leu Phe Leu Leu Ser Phe Val
    Asp Leu Trp Phe Leu Trp Ile Trp Ile Ile Thr Trp Leu
    Glu Val Tyr Trp Val Tyr Val Tyr Val Val Gly Tyr Ile
    Gln Met         Met     Met     Met Met Ala     Met
    His

Gln Gly Ser Leu Asp Lys Arg
Asp Ser Thr Ile Asn Arg Lys
Asn Thr Gly Val Glu
Glu Ala Ala Met Gln
His             His
```

wherein vertically aligned groups of amino acids represent alternative amino acids at the position shown, and positions 20 and 21 of sequence (a) are not Gly and Val respectively, or a polypeptide which is up to four amino acids shorter or longer

provided (i) that the C-terminal Lys-Arg, Arg-Lys, Lys-Lys or Arg-Arg entity is maintained, (ii) that there is a positively charged residue within 5 residues of the N-terminus, (iii) that there is a generally hydrophobic region at or adjacent the middle of the sequence,

and (iv) that the three amino acids N-terminal to the C-terminal Lys/Arg, Lys/Arg entity are maintained.

2. A polypeptide according to claim 1 having an amino acid sequence as follows:

(a)

```
1                                                   10
Met Lys Trp Val Ser Phe Ile Ser Leu Leu Phe Leu Phe Ser
    Arg Phe Leu Thr Trp Leu Thr Ile Ile Trp Ile Trp Thr
    His Tyr Ile Gly Tyr Val Gly Val Val Tyr Val Tyr Gly
    Gln     Met Ala     Met Ala Met Met     Met     Ala
    Asn

                        20
Ser Ala Tyr Ser Arg Ser Leu Asp Lys Arg
Thr Thr Phe Thr Lys Thr Ile Glu Arg Lys
Gly Gly Trp Gly His Gly Val Asn
Ala Ser     Ala Gln Ala Met Gln
                Asn         His
```

or

(b)

```
Met Asn Ile Phe Tyr Ile Phe Leu Phe Leu Leu Ser Phe Val
    Asp Leu Trp Phe Leu Trp Ile Trp Ile Ile Thr Trp Leu
    Glu Val Tyr Trp Val Tyr Val Tyr Val Val Gly Tyr Ile
    Gln Met         Met     Met     Met Met Ala     Met
    His

Gln Gly Ser Leu Asp Lys Arg
Asp Ser Thr Ile Asn Arg Lys
Asn Thr Gly Val Glu
Glu Ala Ala Met Gln
His             His
```

wherein vertically aligned groups of amino acids represent alternative amino acids at the position shown, or a polypeptide which is up to four amino acids longer

provided that positions 20 and 21 of sequence (a) are not Gly and Val respectively.

3. A polypeptide according to Claim I or 2 and having an amino acid sequence:

(a) H_2N-Met-Lys-Trp-Val-Ser-Phe-Ile-Ser-Leu-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-COOH

or

(b) H_2N-Met-Asn-Ile-Phe-Tyr-Ile-Phe-Leu-Phe-Leu-Leu-Ser-Phe-Val-Gln-Gly-Ser-Leu-Asp-Lys-Arg-COOH

4. A fusion compound comprising a polypeptide according to Claim 1, 2 or 3 linked at the carboxyl terminal to the N-terminal residue of a second polypeptide.

5. A fusion compound according to Claim 4 wherein the polypeptide of Claim 1 or 2 is linked directly to said second polypeptide.

6. A fusion compound according to Claim 5 wherein the second polypeptide is a naturally-occurring human serum albumin (HSA), a modified human serum albumin having a tertiary structure similar to that of HSA and having at least one physiological function in common with HSA, or a fragment of either.

7. A polynucleotide sequence encoding a polypeptide according to Claim 1, 2 or 3 or a fusion compound according to Claim 4.

8. A polynucleotide according to Claim 7 selected from the sequences shown in Tables 2 and 3.

9. A DNA construct comprising a suitable control region or regions and a polynucleotide according to Claim 7 or 8, said polynucleotide being under the control of the control region.

10. A host transformed with a DNA construct according to Claim 9.

11. Saccharomyces cerevisiae or Schizosaccharomyces pombe according to Claim 10.

12. A process for preparing a polypeptide, comprising cultivating a host according to Claim 10 or 11 and obtaining therefrom the polypeptide expressed by the said polynucleotide or a modified version of the polypeptide.

Patentansprüche

1. Polypeptid mit einer Aminosäuresequenz wie folgt:

(a)

```
1                                                   10
Met Lys Trp Val Ser Phe Ile Ser Leu Leu Phe Leu Phe Ser
    Arg Phe Leu Thr Trp Leu Thr Ile Ile Trp Ile Trp Thr
    His Tyr Ile Gly Tyr Val Gly Val Val Tyr Val Tyr Gly
    Gln     Met Ala     Met Ala Met Met     Met     Ala
    Asn

                        20
Ser Ala Tyr Ser Arg Ser Leu Asp Lys Arg
Thr Thr Phe Thr Lys Thr Ile Glu Arg Lys
Gly Gly Trp Gly His Gly Val Asn
Ala Ser     Ala Gln Ala Met Gln
                Asn         His
```

oder

(b)

```
Met Asn Ile Phe Tyr Ile Phe Leu Phe Leu Leu Ser Phe Val
    Asp Leu Trp Phe Leu Trp Ile Trp Ile Ile Thr Trp Leu
    Glu Val Tyr Trp Val Tyr Val Tyr Val Val Gly Tyr Ile
    Gln Met         Met     Met     Met Met Ala     Met
    His

                Gln Gly Ser Leu Asp Lys Arg
                Asp Ser Thr Ile Asn Arg Lys
                Asn Thr Gly Val Glu
                Glu Ala Ala Met Gln
                His             His
```

wobei vertikal ausgerichtete Gruppen von Aminosäuren für alternative Aminosäuren an der angegebenen Stelle stehen und die Stellen 20 bzw. 21 der Sequenz (a) nicht von Gly bzw. Val eingenommen werden, oder ein Polypeptid, welches bis zu vier Aminosäuren kürzer oder länger ist, wobei gilt, daß

(i) die C-terminale Lys-Arg-, Arg-Lys-, Lys-Lys- oder Arg-Arg-Einheit erhalten ist,
(ii) es einen positiv geladenen Rest innerhalb von fünf Resten des N-Terminus gibt,
(iii) ein im allgemeinen hydrophober Bereich in oder nächst der Mitte der Sequenz vorhanden ist und
(iv) die drei Aminosäuren N-terminal zu der C-terminalen Lys/Arg, Lys/Arg-Einheit erhalten sind.

2. Polypeptid nach Anspruch 1 mit einer Aminosäuresequenz wie folgt:

(a)

```
1                                                  10
Met Lys Trp Val Ser Phe Ile Ser Leu Leu Phe Leu Phe Ser
    Arg Phe Leu Thr Trp Leu Thr Ile Ile Trp Ile Trp Thr
    His Tyr Ile Gly Tyr Val Gly Val Val Tyr Val Tyr Gly
    Gln     Met Ala     Met Ala Met Met     Met     Ala
    Asn

        Ser Ala Tyr Ser Arg Ser Leu Asp Lys Arg
        Thr Thr Phe Thr Lys Thr Ile Glu Arg Lys
        Gly Gly Trp Gly His Gly Val Asn
        Ala Ser     Ala Gln Ala Met Gln
                        Asn         His
```

oder

(b)

```
Met Asn Ile Phe Tyr Ile Phe Leu Phe Leu Leu Ser Phe Val
    Asp Leu Trp Phe Leu Trp Ile Trp Ile Ile Thr Trp Leu
    Glu Val Tyr Trp Val Tyr Val Tyr Val Val Gly Tyr Ile
    Gln Met         Met     Met     Met Met Ala     Met
    His.

    Gln Gly Ser Leu Asp Lys Arg
    Asp Ser Thr Ile Asn Arg Lys
    Asn Thr Gly Val Glu
    Glu Ala Ala Met Gln
    His             His
```

wobei vertikal ausgerichtete Gruppen von Aminosäuren für alternative Aminosäuren an der angegebenen Stelle stehen, oder ein Polypeptid, welches bis zu vier Aminosäuren länger ist, wobei gilt, daß die Stellungen 20 bzw. 21 der. Sequenz (a) nicht von Gly bzw. Val eingenommen sind.

3. Polypeptid nach Anspruch 1 oder 2 mit einer Aminosäuresequenz:

(a) H_2N-Met-Lys-Trp-Val-Ser-Phe-Ile-Ser-Leu-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-COOH

oder

(b) H_2N-Met-Asn-Ile-Phe-Tyr-Ile-Phe-Leu-Phe-Leu-Leu-Ser-Phe-Val-Gln-Gly-Ser-Leu-Asp-Lys-Arg-COOH

4. Fusionsverbindung, umfassend ein Polypeptid nach Anspruch 1, 2 oder 3, welches am Carboxylende an den N-terminalen Rest eines zweiten Polypeptids gebunden ist.

5. Fusionsverbindung nach Anspruch 4, wobei das Polypeptid nach Anspruch 1 oder 2 direkt an das zweite Polypeptid gebunden ist.

6. Fusionsverbindung nach Anspruch 5, wobei das zweite Polypeptid aus natürlich vorkommendem Humanserumalbumin (HSA), einem modifizierten Humanserumalbumin mit tertiärer Struktur ähnlich derjenigen von HSA und mindestens einer physiologischen Funktion, die mit HSA gemeinsam ist, oder einem Fragment eines derselben besteht.

7. Polynucleotidsequenz mit Codierung für ein Polypeptid nach Anspruch 1, 2 oder 3 oder eine Fusionsverbindung nach Anspruch 4.

8. Polynucleotid nach Anspruch 7, ausgewählt aus den in Tabellen 2 und 3 dargestellten Sequenzen.

9. DNA-Konstrukt, umfassend (einen) geeignete(n) Steuerbereich oder -bereiche und ein Polynucleotid nach Anspruch 7 oder 8, wobei das Polynucleotid unter der Kontrolle des Steuerbereichs steht.

10. Mit einem DNA-Konstrukt nach Anspruch 9 transformierter Wirt.

11. Saccharomyces cerevisiae oder Schizosaccharomyces pombe nach Anspruch 10.

12. Verfahren zur Herstellung eines Polypeptids durch Züchten eines Wirts nach Anspruch 10 oder 11 und Gewinnen des durch das Polynucleotid exprimierten Polypeptids oder einer modifizierten Version des Polypeptids aus diesem.

Revendications

1. Polypeptide ayant une séquence d'acides aminés comme suit :

(a)

```
1                                                  10
Met Lys Trp Val Ser Phe Ile Ser Leu Leu Phe Leu Phe Ser
    Arg Phe Leu Thr Trp Leu Thr Ile Ile Trp Ile Trp Thr
    His Tyr Ile Gly Tyr Val Gly Val Val Tyr Val Tyr Gly
    Gln     Met Ala     Met Ala Met Met     Met     Ala
    Asn

                        20
Ser Ala Tyr Ser Arg Ser Leu Asp Lys Arg
Thr Thr Phe Thr Lys Thr Ile Glu Arg Lys
Gly Gly Trp Gly His Gly Val Asn
Ala Ser     Ala Gln Ala Met Gln
                Asn         His
```

ou

(b)

```
Met Asn Ile Phe Tyr Ile Phe Leu Phe Leu Leu Ser Phe Val
    Asp Leu Trp Phe Leu Trp Ile Trp Ile Ile Thr Trp Leu
    Glu Val Tyr Trp Val Tyr Val Tyr Val Val Gly Tyr Ile
    Gln Met         Met     Met     Met Met Ala     Met
    His

Gln Gly Ser Leu Asp Lys Arg
Asp Ser Thr Ile Asn Arg Lys
Asn Thr Gly Val Glu
Glu Ala Ala Met Gln
His             His
```

dans laquelle les groupes d'acides aminés alignés à la verticale représentent des acides aminés alternatifs à la position indiquée et les positions 20 et 21 de la séquence (a) ne sont ni Gly ni Val, respectivement, ou un polypeptide qui compte jusqu'à quatre acides aminés de moins ou de plus, à condition (i) que l'entité C-terminale Lys-Arg, Arg-Lys, Lys-Lys ou Arg-Arg soit maintenue, (ii) qu'il y ait un résidu chargé positivement parmi les 5 résidus de l'extrémité N-terminale, (iii) qu'il y ait une région généralement hydrophobe au milieu de la séquence ou dans une position adjacente, et (iv) que les trois acides aminés N-terminaux jusqu'à l'entité C-terminale Lys/Arg, Lys/Arg soit maintenus.

2. Polypeptide selon la revendication 1 ayant une séquence d'acides aminés comme suit :

(a)

```
1                                       10
Met Lys Trp Val Ser Phe Ile Ser Leu Leu Phe Leu Phe Ser
    Arg Phe Leu Thr Trp Leu Thr Ile Ile Trp Ile Trp Thr
    His Tyr Ile Gly Tyr Val Gly Val Val Tyr Val Tyr Gly
    Gln     Met Ala     Met Ala Met Met     Met     Ala
    Asn

                        20
Ser Ala Tyr Ser Arg Ser Leu Asp Lys Arg
Thr Thr Phe Thr Lys Thr Ile Glu Arg Lys
Gly Gly Trp Gly His Gly Val Asn
Ala Ser     Ala Gln Ala Met Gln
                Asn         His
```

ou

(b)

```
Met Asn Ile Phe Tyr Ile Phe Leu Phe Leu Leu Ser Phe Val
    Asp Leu Trp Phe Leu Trp Ile Trp Ile Ile Thr Trp Leu
    Glu Val Tyr Trp Val Tyr Val Tyr Val Val Gly Tyr Ile
    Gln Met         Met     Met     Met Met Ala     Met
    His

    Gln Gly Ser Leu Asp Lys Arg
    Asp Ser Thr Ile Asn Arg Lys
    Asn Thr Gly Val Glu
    Glu Ala Ala Met Gln
    His             His
```

dans laquelle les groupes d'acides aminés alignés à la verticale représentent des acides aminés alternatifs à la position indiquée, ou un polypeptide qui compte jusqu'à quatre acides aminés de plus, à condition que les positions 20 et 21 de la séquence (a) ne soient ni Gly ni Val, respectivement.

3. Polypeptide selon la revendication 1 ou 2 et ayant une séquence d'acides aminés :

(a) H_2N-Met-Lys-Trp-Val-Ser-Phe-Ile-Ser-Leu-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-COOH

ou

(b) H_2N-Met-Asn-Ile-Phe-Tyr-Ile-Phe-Leu-Phe-Leu-Leu-Ser-Phe-Val-Gln-Gly-Ser-Leu-Asp-Lys-Arg-COOH

4. Composé de fusion comprenant un polypeptide selon la revendication 1, 2 ou 3 lié au niveau de l'extrémité carboxyle au résidu N-terminal d'un deuxième polypeptide.

5. Composé de fusion selon la revendication 4, dans lequel le polypeptide selon la revendication 1 ou 2 est directement lié audit deuxième polypeptide.

6. Composé de fusion selon la revendication 5, dans lequel le deuxième polypeptide est une sérum-albumine humaine naturelle (SAH), une sérum-albumine humaine modifiée ayant une structure tertiaire similaire à celle de la SAH et ayant au moins une fonction physiologique en commun avec la SAH, ou un fragment de l'une ou l'autre.

7. Séquence polynucléotidique codant un polypeptide selon la revendication 1, 2 ou 3 ou composé de fusion selon la revendication 4.

8. Polynucléotide selon la revendication 7 choisi parmi les séquences indiquées dans les Tableaux 2 et 3.

9. Fragment d'ADN comprenant une ou des régions de commande appropriées et un polynucléotide selon la revendication 7 ou 8, ledit polynucléotide étant sous la commande de la région de commande.

10. Hôte transformé avec un fragment d'ADN selon la revendication 9.

11. Saccharomyces cerevisiae ou Schizosaccharomyces pombe selon la revendication 10.

12. Procédé de préparation d'un polypeptide, consistant à cultiver un hôte selon la revendication 10 ou 11 et à obtenir de celui-ci le polypeptide exprimé par ledit polynucléotide ou une version modifiée du polypeptide.

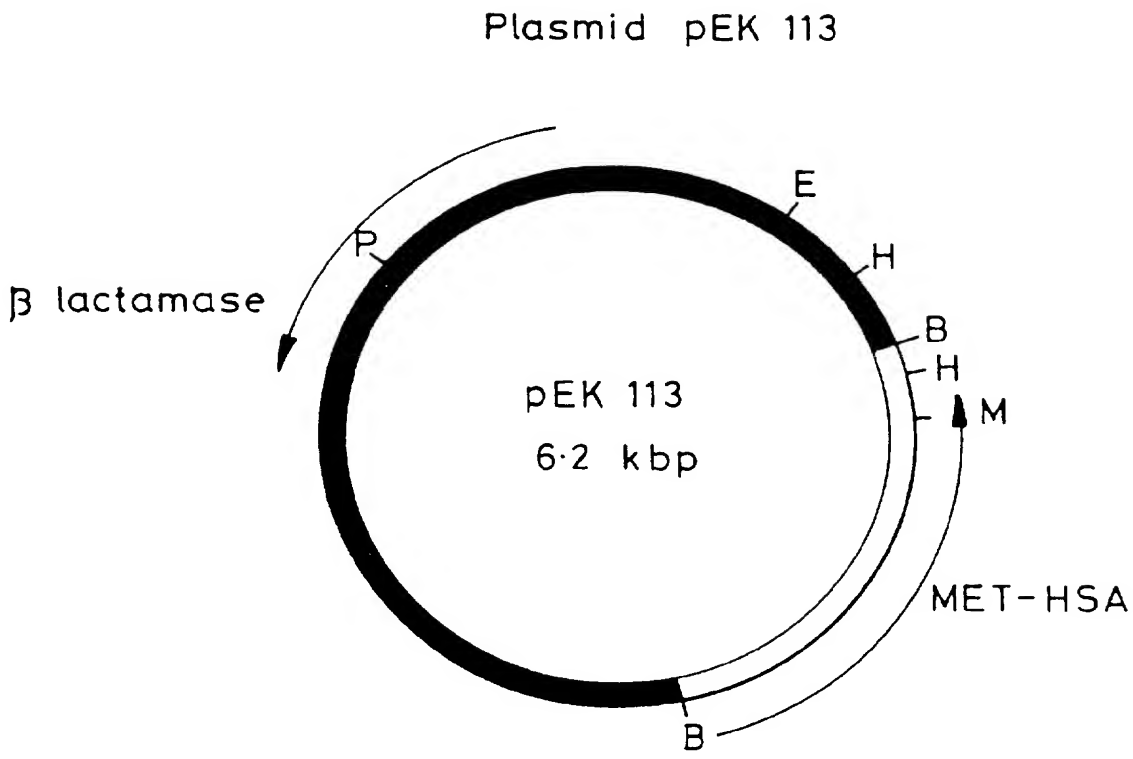
Plasmid pEK 113
β lactamase
P
E
H
B
H
M
pEK 113
6·2 kbp
MET-HSA
B
pBR 322 DNA
MET-HSA DNA
Restriction endonuclease sites

Fig. 1

Plasmid pEK 25

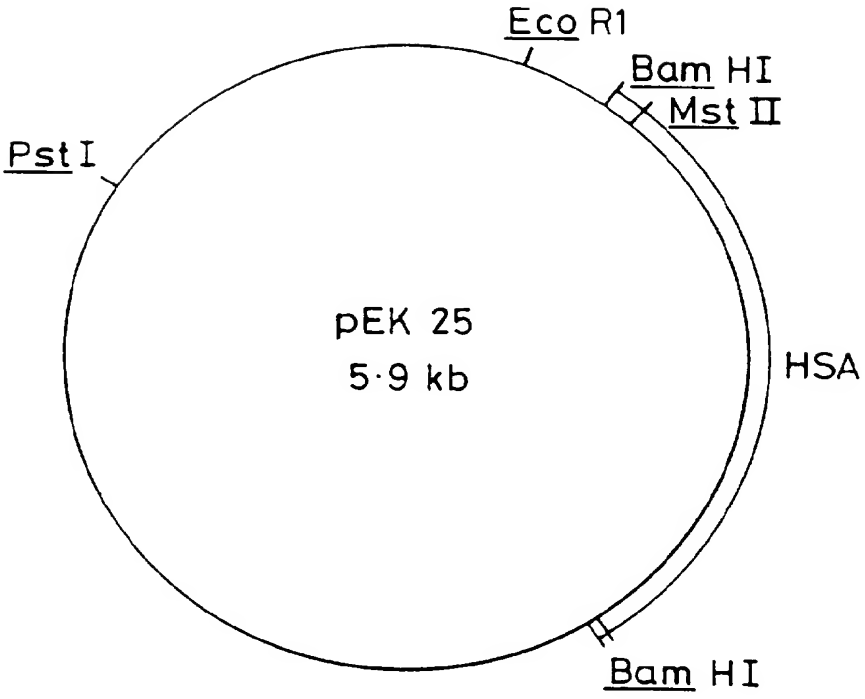

Fig. 2

Plasmid pAYE 230

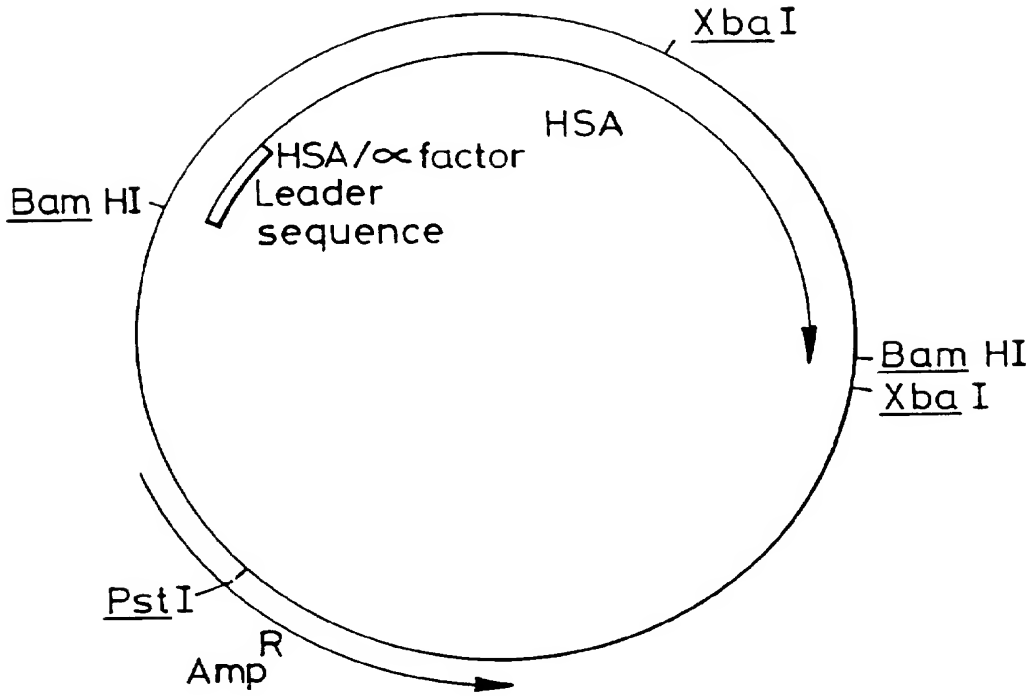

Fig. 3

Plasmid pAYE 238

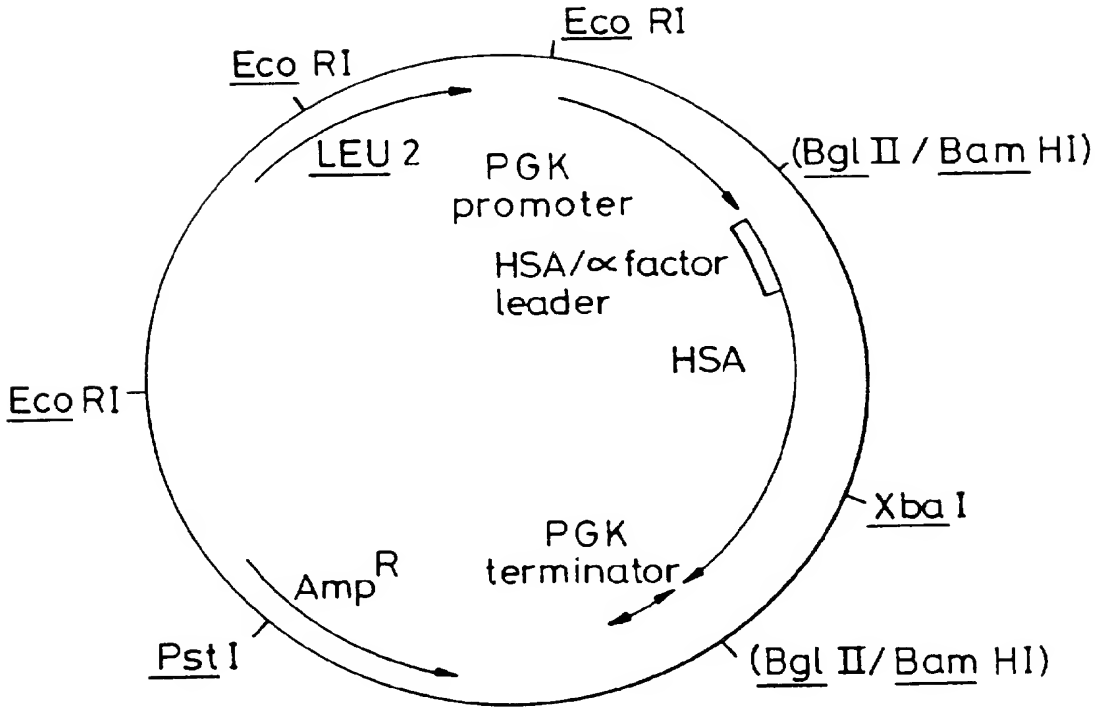

Fig. 4

Plasmid pAYE 304

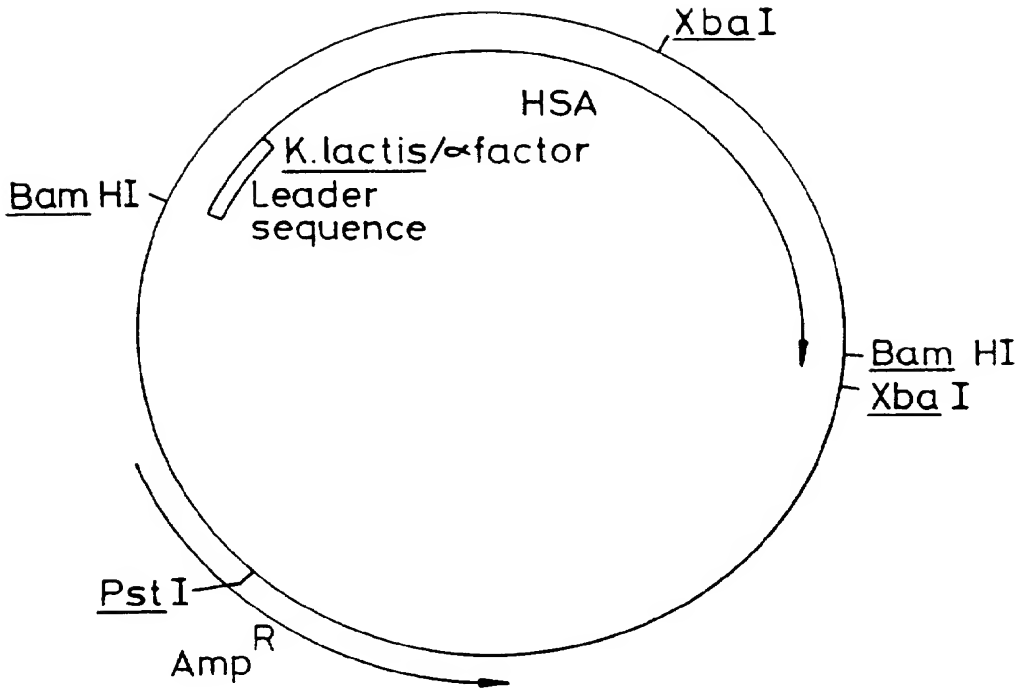

Fig. 5

Plasmid pAYE 305

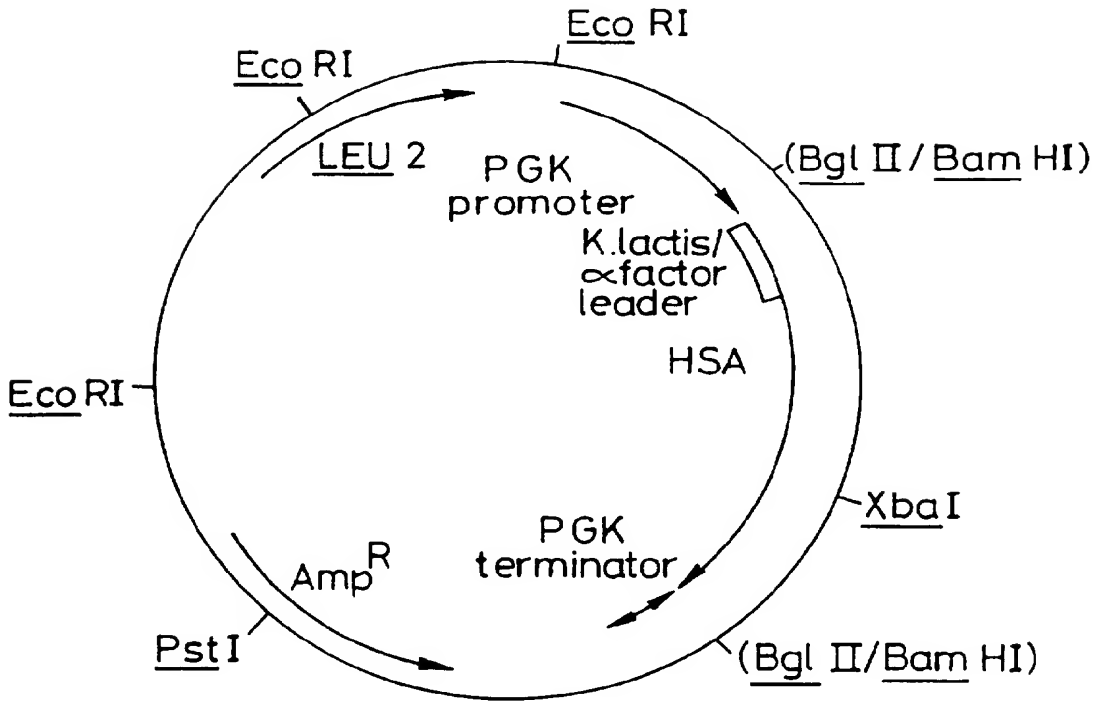

Fig. 6